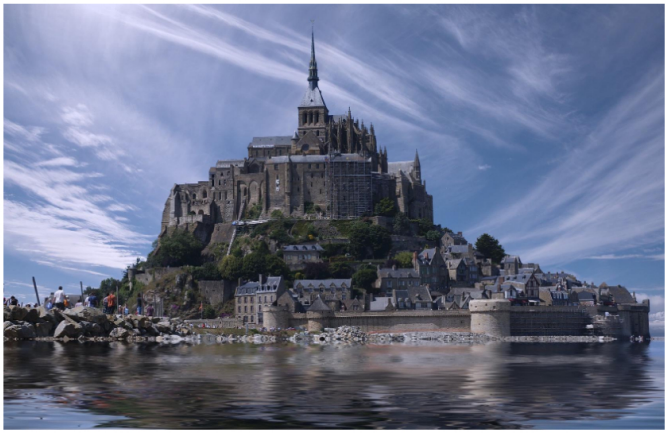
Congratulations! You passed!

Grade received 80% | Latest Submission Grade 80% | To pass 80% or higher

Go to next item

1. This example is adapted from a real production application, but with details disguised to protect confidentiality. 1 / 1 point



You are a famous researcher in the City of Peacetopia. The people of Peacetopia have a common characteristic: they are afraid of birds. To save them, you have **to build an algorithm that will detect any bird flying over Peacetopia** and alert the population.

The City Council gives you a dataset of 10,000,000 images of the sky above Peacetopia, taken from the city's security cameras. They are labeled:

- y = 0: There is no bird on the image
- y = 1: There is a bird on the image

Your goal is to build an algorithm able to classify new images taken by security cameras from Peacetopia.

There are a lot of decisions to make:

- What is the evaluation metric?
- How do you structure your data into train/dev/test sets?

Metric of success

The City Council tells you the following that they want an algorithm that

1. Has high accuracy.
2. Runs quickly and takes only a short time to classify a new image.
3. Can fit in a small amount of memory, so that it can run in a small processor that the city will attach to many different security cameras.

You meet with them and ask for just one evaluation metric. True/False?

- False
- True: (selected)

Expand

Correct
Yes. The goal is to have one metric that focuses the development effort and increases iteration velocity.

2. The city asks for your help in further defining the criteria for accuracy, runtime, and memory. How would you suggest they identify the criteria? 1 / 1 point

- Suggest to them that they define which criterion is most important. Then, set thresholds for the other two. (selected)
- Suggest that they purchase more infrastructure to ensure the model runs quickly and accurately.
- Suggest to them that they focus on whichever criterion is important and then eliminate the other two.

Expand

Correct

Yes. The thresholds provide a way to evaluate models head to head.

3. Based on the city's requests, which of the following would you say is true? **1 / 1 point**

- (•) Accuracy is an optimizing metric; running time and memory size are satisfying metrics.
- ( ) Accuracy, running time and memory size are all satisfying metrics because you have to do sufficiently well on all three for your system to be acceptable.
- ( ) Accuracy is a satisfying metric; running time and memory size are an optimizing metric.
- ( ) Accuracy, running time and memory size are all optimizing metrics because you want to do well on all three.

Expand

Correct

4. You propose a 95/2.5%/2.5% for train/dev/test splits to the City Council. They ask for your reasoning. Which of the following best justifies your proposal? **1 / 1 point**

- ( ) The most important goal is achieving the highest accuracy, and that can be done by allocating the maximum amount of data to the training set.
- (•) With a dataset comprising 10M individual samples, 2.5% represents 250k samples, which should be more than enough for dev and testing to evaluate bias and variance.
- ( ) The emphasis on the training set will allow us to iterate faster.
- ( ) The emphasis on the training set provides the most accurate model, supporting the memory and processing satisficing metrics.

Expand

Correct
Yes. The purpose of dev and test sets is fulfilled even with smaller percentages of the data.

5. Now that you've set up your train/dev/test sets, the City Council comes across another 1,000,000 images from social media and offers them to you. These images are different from the distribution of images the City Council had originally given you, but you think it could help your algorithm. You should add the citizens' data to the training set. True/False? **0 / 1 point**

- (•) False
- ( ) True

Expand

Incorrect
No. Adding this data to the training set will change the training set distribution. However, it is not a problem to have different training and dev distributions. In contrast, it would be very problematic to have different dev and test set distributions.

6. One member of the City Council knows a little about machine learning and thinks you should add the 1,000,000 citizens' data images to the dev set. You object because: (Choose all that apply) **1 / 1 point**

- [x] The dev set no longer reflects the distribution of data (security cameras) you most care about.

  Correct
  Yes. The performance of the model should be evaluated on the same distribution of images it will see in production.

- [ ] The 1,000,000 citizens' data images do not have a consistent x-->y mapping as the rest of the data.
- [x] This would cause the dev and test set distributions to become different. This is a bad idea because you're not aiming where you want to hit.

  Correct
  Yes. Adding a different distribution to the dev set will skew bias.

- [ ] A bigger test set will slow down the speed of iterating because of the computational expense of evaluating models on the test set.

Expand

Correct
Great, you got all the right answers.

7. Human performance for identifying birds is < 1%, training set error is 5.2% and dev set error is 7.3%. Which of the options below is the best next step? **1 / 1 point**

- Get more data or apply regularization to reduce variance.
- Try an ensemble model to reduce bias and variance.
- Train a bigger network to drive down the >4.0% training error.
- Validate the human data set with a sample of your data to ensure the images are of sufficient quality.

Expand

Correct
Yes. Avoidable bias is >4.2% which is larger than the 2.1% variance.

8. You want to define what human-level performance is to the city council. Which of the following is the best answer? **0 / 1 point**

- The average performance of all their ornithologists (0.5%).
- The average of all the numbers above (0.66%).
- The performance of their best ornithologist (0.3%).
- The average of regular citizens of Peacetopia (1.2%).

Expand

Incorrect
No. The average reflects a range of skills, not the best.

9. A learning algorithm's performance can be better than human-level performance but it can never be better than Bayes error. True/False? **1 / 1 point**

- False.
- True.

Expand

Correct
Yes. By definition, human level error is worse than Bayes error.

10. You find that a team of ornithologists debating and discussing an image gets an even better 0.1% performance, so you define that as "human-level performance." After working further on your algorithm, you end up with the following: **1 / 1 point**

Human-level performance	0.1%
Training set error	2.0%
Dev set error	2.1%

Based on the evidence you have, which two of the following four options seem the most promising to try? (Check two options.)

- [x] Try decreasing regularization.

Correct

- [ ] Try increasing regularization.
- [ ] Get a bigger training set to reduce variance.
- [x] Train a bigger model to try to do better on the training set.

Correct

Correct

Expand

Correct
Great, you got all the right answers.

11. You also evaluate your model on the test set, and find the following: 1 / 1 point

Human-level performance	0.1%
Training set error	2.0%
Dev set error	2.1%
Test set error	7.0%

What does this mean? (Check the two best options.)

- [x] You have overfit to the dev set.

Correct

- [ ] You have underfitted to the dev set.
- [ ] You should get a bigger test set.
- [x] You should try to get a bigger dev set.

Correct

Expand

Correct
Great, you got all the right answers.

12. After working on this project for a year, you finally achieve: Human-level performance, 0.10%, Training set error, 0.05%, Dev set error, 0.05%. Which of the following are likely? (Check all that apply.) 1 / 1 point

- [ ] There is still avoidable bias.
- [x] The model has recognized emergent features that humans cannot. (Chess and Go for example)

Correct
Yes. When Google beat the world Go champion, it was recognized that it was making deeper moves than humans.

- [ ] This result is not possible since it should not be possible to surpass human-level performance.
- [x] Pushing to even higher accuracy will be slow because you will not be able to easily identify sources of bias.

Correct
Yes. Exceeding human performance means you are close to Bayes error.

Expand

Correct
Great, you got all the right answers.

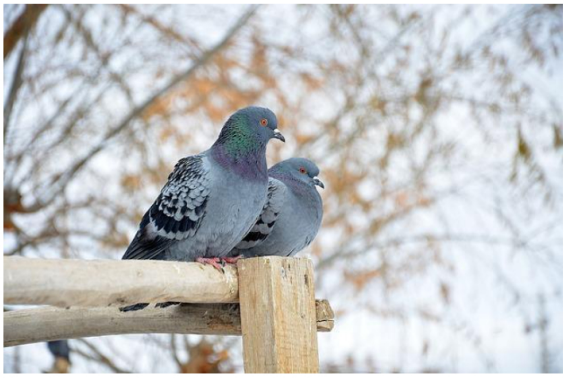
13. It turns out Peacetopia has hired one of your competitors to build a system as well. Your system and your competitor both deliver systems with about the same running time and memory size. However, your system has higher accuracy! However, when Peacetopia tries out your and your competitor's systems, they conclude they actually like your competitor's system better, because even though you have higher overall accuracy, you have more false negatives (failing to raise an alarm when a bird is in the air). What should you do? 1 / 1 point

- ( ) Ask your team to take into account both accuracy and false negative rate during development.
- ( ) Pick false negative rate as the new metric, and use this new metric to drive all further development.
- ( ) Look at all the models you've developed during the development process and find the one with the lowest false negative error rate.
- (•) Rethink the appropriate metric for this task, and ask your team to tune to the new metric.

Expand

Correct

14. You've handily beaten your competitor, and your system is now deployed in Peacetopia and is protecting the citizens from birds! But over the last few months, a new species of bird has been slowly migrating into the area, so the performance of your system slowly degrades because your data is being tested on a new type of data. 1 / 1 point



You have only 1,000 images of the new species of bird. The city expects a better system from you within the next 3 months. Which of these should you do first?

- Put the 1,000 images into the training set so as to try to do better on these birds.
- Add the 1,000 images into your dataset and reshuffle into a new train/dev/test split.
- Try data augmentation/data synthesis to get more images of the new type of bird.
- Use the data you have to define a new evaluation metric (using a new dev/test set) taking into account the new species, and use that to drive further progress for your team.

Expand

Correct

15. The City Council thinks that having more cats in the city would help scare off birds. They are so happy with your work on the Bird detector that they also hire you to build a Cat detector. You have a huge dataset of 100,000,000 cat images. Training on this data takes about two weeks. Which of the statements do you agree with? (Check all that agree.) 0 / 1 point

- [x] You could consider a tradeoff where you use a subset of the cat data to find reasonable performance with reasonable iteration pacing.

  Correct
  Yes. This is similar to satisficing metrics where "good enough" determines the size of the data.

- [x] Given a significant budget for cloud GPUs, you could mitigate the training time.

  Correct
  Yes. More resources will allow you to iterate faster.

- [ ] Accuracy should exceed the City Council's requirements but the project may take as long as the bird detector because of the two week training/iteration time.
- [ ] With the experience gained from the Bird detector you are confident to build a good Cat detector on the first try.

Expand

Incorrect
You didn't select all the correct answers